

H17/A12 人間の知覚特性を考慮したマルチモーダル 情報処理システムに関する研究(1節 共同プロジェ クト研究の理念と概要, 第4章 共同プロジェクト研 究)

雑誌名	東北大学電気通信研究所研究活動報告
号	13
ページ	149-151
発行年	2006
URL	http://hdl.handle.net/10097/40644

課題番号 H17/A12

採択回数	1	2	3
------	---	---	---

人間の知覚特性を考慮したマルチモーダル情報処理システムに関する研究

[1] 組織

代表者：田中 章浩

(東京大学大学院人文社会系研究科)

対応者：鈴木 陽一

(東北大学電気通信研究所)

分担者：坂本 修一

(東北大学電気通信研究所)

研究費：物件費 29 万 2 千円，旅費 23 万 4 千円

[2] 研究経過

(本プロジェクトの目的・概要，及び，研究会，研究打ち合わせ・研究討論会，研究発表会，研究集会等の開催状況を記載して下さい。)

コミュニケーションに関わる通信システムでは，情報の発信と受容の担い手は人間であり，人間の知覚特性を十分に考慮したシステムの構築が望まれる。しかし，テレビ番組のように音声を伝達する通信システムは，とりわけ高齢者にとって，情報の受容（聴き取り）が困難であることが多い。このような状況に対応して，発話速度を下げることによって，聴き取りやすい音声を提示する話速変換技術の開発が進められている。

また，話し言葉の理解は一般に聴覚の働きであると考えられがちであるが，話し手の顔が見える状況では，音声情報に加えて，口の動きなどの視覚情報

も非常に重要である。しかし，従来の話速変換システムは視覚情報の併用についてほとんど考慮してこなかった。

そこで本プロジェクトでは，話速を下げて，かつ視覚情報も利用可能なマルチモーダル情報処理システムについての基礎的検討をおこなうことを目的として研究を行った。

以下，研究活動状況の概要を記す。

11/20 研究打ち合わせ

昨年度実施した若年者での実験結果と，本年度実施した高齢者での実験結果について比較検討する議論をおこなった。

2/15 研究打ち合わせ

今年度のこれまでの成果を踏まえて，次年度の研究計画について議論した。

2/22-24 研究発表会

今年度の成果全体についての議論をおこなうとともに，関連する研究プロジェクトの成果も含めて意見交換をおこなった。

[3] 成果

(3-1) 研究成果

昨年度は，高齢者への適用の前段階として，若年者を対象とした実験を実施した。

昨年度得られた成果を実際に高齢者に適用するためには，さまざまな問題を解決する必要がある。例

えば、高齢者は聴力が低く、かつ個人間のばらつきも大きい。また、音声の聴取能力にも個人差がある。そこで本年度は、高齢者を対象とした実験を実施した。

実験で用いる言語刺激には、数的に限定された刺激単語を用いた。統制された単語を厳選することで、より詳細に映像の効果について分析するためである。具体的には、三省堂「NTT データベース 日本語の語彙特性」から、1 子音違いで口の動きも異なる 4 モーラ単語のペアを選定した。各モーラ位置において 1 子音違いの単語を 5 ペアずつ用いた（第 1 モーラ位置でのペア単語の例として「トウハツ」と「モウハツ」）。女性話者が発話した単語の音声と映像を DV に記録した。実験では、同じ速度の音声と映像の提示タイミングをずらす非同期 AV 条件、音声を時間伸長し、映像と音声の速度が異なる伸長 AV 条件、時間伸長音声のみを提示する伸長 A 条件、統制条件（原音声のみ、映像のみ）を実施した。音声を聴き取りにくくするため、音声にノイズを重畳して提示した。ただし、若年者と同じ大きさのノイズを重畳すると、高齢者にとっては音声聴取が困難になる。そこで、予備実験によって適切なノイズの大きさを決定した。単語理解度はモーラ位置別に算出した。各条件、モーラ位置別の単語理解度をもとに、映像の効果（AV benefit）を算出した。

実験の結果、単語の語頭から語中では映像と音声の時間差が 200 ms 程度までであれば問題なく読唇が可能であるが、時間差が 400 ms 程度になると効率よく読唇できなくなることがわかった。

昨年若年者の結果では、視聴覚間に 400 ms 程度の時間差が存在しても、視聴覚間の時間差が一定であれば語末では視聴覚が同期しているときと同等に読唇が可能であった。これは語頭から語末という短時間に視聴覚間の同期について補正が生じたためであると考えられる。しかし、高齢者では視聴覚間

に 400 ms 程度の時間差が存在した場合、視聴覚が同期しているときと比べると成績が低下した。つまり、視聴覚間の補正は高齢者では生じなかった。また、若齢者と同様、高齢者においても、映像と音声の速度が異なり、視聴覚間の時間差が一定でない状況下では、補正が生じないことも示唆された。

また、話速変換による正の効果が一部の条件において確認された。話速変換による正の効果は単語では得られにくく、文・文章を刺激とした研究において広く報告されている。したがって、今後、実験刺激として文・文章を用いた場合には、映像の効果に加えて話速変換の効果もより顕著に得られることが期待できる。

このような視聴覚音声認知の時間特性に関する検討を進めることによって、音声の時間伸長の効果と読唇による映像の効果を両方活用できるような音声聴取補助システムの構築へとつながることが期待される。

（3-2）波及効果と発展性など

（大型プロジェクトへの発展・国際会議（シンポジウム）への発展・学外研究者との交流、共同研究による効果・研究者ネットワークの拡大・若手研究者の育成・新研究領域の開拓・成果の他分野への応用・萌芽的研究への発展等）

本プロジェクトは、電気通信分野と心理学分野の研究者により構成されている。開始して 2 年の間に、電気通信分野と心理学分野の研究者間での交流の場を継続的に得ることができた。電気通信分野の学会（日本音響学会、電子情報通信学会）のみならず、心理学分野の学会（日本認知心理学会、日本基礎心理学会）で成果を発表したところ、学術的に高い評価を得て、日本認知心理学会優秀発表賞（社会的貢献度部門）を受賞した。このように分野間の交流も

飛躍的に活性化し、学際的なプロジェクトに発展しつつある。

〔4〕 成果資料（以下10.5ポイント）

（3頁以降は、本プロジェクトで研究された研究成果が掲載されている主要論文リストを、15件程度を掲載してください。）

【平成17年度】

（1）話速変換による話者映像のずれが単語理解度に及ぼす影響

津村光美, 田中章浩, 坂本修一, 鈴木陽一
日本音響学会講演論文集, 2-Q-3 (2005 年 9 月), 473-474.

（2）話速変換単語音声聴取に非同期話者映像がもたらす影響

津村光美, 田中章浩, 坂本修一, 鈴木陽一
電子情報通信学会技術研究報告, HIP2005-98 (2005 年 12 月), 103-108.

（3）単語音声と話者映像のずれが視聴覚統合に及ぼす影響

津村光美, 田中章浩, 坂本修一, 鈴木陽一
日本音響学会講演論文集, 3-P-9 (2006 年 3 月).

（4）映像と音声のタイミングおよび速度の違いが単語音声認知に及ぼす影響

田中章浩, 津村光美, 坂本修一, 鈴木陽一
Technical Report on Attention and Cognition, 2006, No.13.

【平成18年度】

（5）映像と音声の提示タイミングおよび提示速度の違いが単語音声認知に及ぼす影響

田中章浩・津村光美・坂本修一・鈴木陽一
日本基礎心理学会第25回大会, 広島国際大学, 2006 年.

（6）視聴覚の提示タイミングおよび提示速度の違いが単語音声認知に及ぼす影響

田中章浩・坂本修一・津村光美・鈴木陽一
日本音響学会聴覚研究会資料, 36(4), 369-374, 2006 年.

（7）音声言語理解に発話速度とポーズ長が及ぼす影響

田中章浩・坂本修一・鈴木陽一
日本認知心理学会第4回大会発表論文集, p.16, 2006 年. 【優秀発表賞受賞】

（8）Tanaka, A., Sakamoto, S., Tsumura, K. & Suzuki, Y. "Effects of Intermodal Timing difference and speed difference on auditory-visual speech perception."

Journal of the Acoustical Society of America, Vol.120, No.5, Pt.2, pp.3347-3348, 2006 年.